

### **REMARKS**

Reconsideration and allowance in view of the foregoing amendment and the following remarks are respectfully requested.

#### **Rejection of Claims 22-25, 27, 29-32 and 34 Under 35 U.S.C. §103(a)**

The Office Action rejects claims 22-25, 27, 29-32 and 34 under 35 U.S.C. §103(a) as being unpatentable over Ezzat et al. ("Visual Speech Synthesis by Morphing Visemes") ("Ezzat et al.") in view of Jiang et al. ("Visual Speech Analysis with Application to Mandarin Speech Training") ("Jiang et al.") in view of Hon et al. ("Automatic Generation of Synthesis Unites for Trainable Text-to-Speech Systems") ("Hon et al."). Applicant again traverses this rejection and shall discuss the status of the arguments as well as presenting a new argument which should clearly tip the balance in Applicant's favor to prevent these references from being combined.

Applicants have articulated in depth the standard for under what circumstances it is appropriate to combine two references as being obvious. In response to Applicant's arguments after the Final Rejection, the Examiner issued an Advisory Action. Applicant appreciates the details provided by the Examiner in the Advisory Action which enable us to provide further information and arguments in favor of the patentability of the claims. Applicant notes that the standard of proof is by a preponderance of the evidence. The record thus far includes evidence on both sides. For example, evidence in favor of Applicant includes that Ezzat et al. deals primarily with visual speech synthesis while Jiang et al. focus on extracting information from existing images. Applicant has argued that Jiang et al.'s reference to synthesis is found in a final portion in Section 5 in which as Jiang et al. states that the speech-to-lip movement synthesis techniques exceed the scope of the paper. The Examiner has noted in his favor that the reference does discuss image synthesis. Accordingly, the Advisory Action states that "due to the similarities between the references and motivation provided, the Examiner feels that it

reasonable to expect that one of skill in the art would be motivated to combine the features of each reference to form a combination as stated in the Final Office Action.” This is primarily in reference to the combination of Ezzat et al. and Jiang et al.

Next, the Advisory Action comments on how the Hon et al. reference explains how the Examiner maintains his assertion that Hon et al. teaches the longest possible candidate images being selected (which Applicant shall address later) and how it would be obvious for image samples to be applied to this process of Hon et al. and thus combined with Ezzat et al. (Applicant also thanks the Examiner for clarifying the “Brand” issue and that was a typographical error.)

The fundamental additional argument Applicant will make will be in reference to whether as is asserted in the Advisory Action, one of skill in the art would combine Hon et al. with Ezzat et al. In the Final Office Action on page 4, near the middle, as Applicant understands from the Advisory Action, the sentence referencing Brand should be deleted and is not used to support a rejection of claims. Accordingly, the arguments regarding Hon et al. is primarily that Hon et al. is asserted to teach the claim unit selection process and “suggests the claimed ‘longest possible candidate image sample is selected’” from the teachings on page 296. After, again noting that the reference appears to suggest the concept of using the longest possible candidate from a large database of possible samples, the Final Office Action simply states that it would have been obvious to combine Hon et al. with Ezzat et al. and Jiang et al. and that the advantage would be that unit selection feature selected from a database of a large amount of candidates would produce a optimal concatenation quality.

Applicant asserts that there are fundamental problems with the obviousness analysis in terms of combining Hon et al. with Ezzat et al. Hon et al. focus, in Section 2, on the synthesis unit and introduce the diphone as containing the transitions between two phones in Section 2.1.

Section 2.2 continues to develop the focus of Hon et al. in that “to achieve a more natural voice quality, one must take more contexts into account, going beyond diphones.” (emphasis added) As we shall see, the Hon et al. directly criticize the use of diphones as inadequate for a more natural voice quality. Section 2.3 of Hon et al. teaches that to achieve the rich context modeling while not introducing more junctions, they introduce a decision-tree clustered phone-based unit as their synthesis unit. They construct the inventory of context-dependent phone units from triphones, quinphones (a phone with two immediate left and right contexts), stress-sensitive phones, word-dependent phones, or a combination of the above. In order to reduce the total number of synthesis units down to a manageable number while incorporating more context, they need to utilize clustering decision trees to cluster similar context-dependent phone units together. In Section of 2.3, they state “unlike senones, our phone-based units require no more junctions than diphone-based systems and yet assure consistency within each unit to achieve better concatenation through rich context modeling.” Applicant simply notes that the entire focus of Hon et al. is to automatically generate these synthesis units that are expressly distanced from the use of diphones. In Section 2.1, diphones are criticized as problematic for the naturalness of synthetic speech and that such speech can “sometimes significantly hampered by the context mismatch between certain diphone units.” The suggestive power of the teachings of this reference to one of skill in the art clearly is to discourage the use of diphones.

Inasmuch as the Examiner has asserted that it would be obvious to one of skill in the art for Hon et al. to be combined with Ezzat et al., we must see how Ezzat et al. treats the use of diphones and whether they utilize diphones or highlight the use of diphones. As we shall see, this is exactly the case, thus removing the ability of one of skill in the art to be motivated to combine these references. Section 7 of Ezzat et al. discusses their audio-visual synchronization approach. They use the Festival TTS system which constructs the final audio stream by

concatenating diphones together. They teach that most diphone-based TTS systems record a corpus of about 1600-2500 diphones. They note also in Section 7, right column at the top, that in order to produce a visual speech stream in synchrony with the audio speech stream, a lip-sync module first extracts the duration of each diphone as computed by the audio module and then the lip-sync module creates an intermediate stream called the viseme transition stream that is defined to be the collection of two end-point visemes and the optical flow correspondence between them. The lip-sync module loads the appropriate viseme transitions into the viseme stream by examining the audio diphones.

How does Ezzat et al. explain the quality and success of using diphones? On page 54 near the top, they state that as a final step, each frame is synthesized using the morph algorithm discussed in Section 5.4. "We have found that the use of TTS timing and phonemic information in this manner produces **very good** quality lip synchronization between the audio and the video." (emphasis added) Clearly the approach that has been found to be "very good" of Ezzat et al. is to use diphones in their synthesis and to use specifically the diphone structure to determine the appropriate sequence of viseme transitions as well as managing the rate of transformations. See Summary sentence, page 54, fifth column. Clearly, in Ezzat et al., the use of diphones is fundamental to the entire process of synchronizing the audio with the video. Inasmuch as it is a requirement that diphones are used in Ezzat et al. and furthermore, since they highlight the very good quality of synchronization between the audio and the video, Applicant respectfully submits that these references expressly teach away from their combinations. Revidence in favor of the patentability of the claims outweighs any other evidence which is on the record which may exist in favor of the combination these references.

An additional reason why this evidence tips the scale in Applicant's favor is that the previously cited law regarding where the proposed modification or combination of prior art

would change the principle of operation of the prior art invention being modified, then the references are not sufficient to render the claims *prima facie* obvious. MPEP 2143.01. In this case, clearly the opposite treatment of the use of diphones is stark and would clearly require a fundamental modification of one or both of the references. For example, to utilize the teachings of Hon et al. in Ezzat et al., one would essentially have to abandon the fundamental principles of Hon et al. which discourage the use of diphones. Similarly, to incorporate the teachings of Hon et al. with Ezzat et al., one would have to modify the diphone approach which was already found to be "very good" and completely modify the synchronization technique in order to accommodate a triphone or quinphones etc. type of synthesis unit for the purpose of synchronization of visemes with speech. Accordingly, for these additional reasons, Applicant submits that the arguments against the motivation to combine is almost unassailable and therefore, Applicant submits that these claims are patentable.

**CONCLUSION**

Having addressed all rejections and objections, Applicant respectfully submits that the subject application is in condition for allowance and a Notice to that effect is earnestly solicited. If necessary, the Commissioner for Patents is authorized to charge or credit the **Law Office of Thomas M. Isaacson, LLC, Account No. 50-2960** for any deficiency or overpayment.

Respectfully submitted,

Date: April 13, 2007

By: \_\_\_\_\_

Correspondence Address:

Thomas A. Restaino  
Reg. No. 33,444  
AT&T Corp.  
Room 2A-207  
One AT&T Way  
Bedminster, NJ 07921

Thomas M. Isaacson

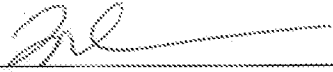
Attorney for Applicant  
Reg. No. 44,166  
Phone: 410-286-9405  
Fax No.: 410-510-1433

**CONCLUSION**

Having addressed all rejections and objections, Applicant respectfully submits that the subject application is in condition for allowance and a Notice to that effect is earnestly solicited. If necessary, the Commissioner for Patents is authorized to charge or credit the **Law Office of Thomas M. Isaacson, LLC, Account No. 50-2960** for any deficiency or overpayment.

Respectfully submitted,

Date: April 13, 2007

By: 

Correspondence Address:

Thomas A. Restaino  
Reg. No. 33,444  
AT&T Corp.  
Room 2A-207  
One AT&T Way  
Bedminster, NJ 07921

Thomas M. Isaacson

Attorney for Applicant  
Reg. No. 44,166  
Phone: 410-286-9405  
Fax No.: 410-510-1433